



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 6, 2026 – 05:39 PM UTC

PDB ID : 8AGH / pdb\_00008agh  
Title : BK Polyomavirus VP1 mutant E73A  
Authors : Sorin, M.N.; Di Maio, A.; Silva, L.M.; Ebert, D.; Delannoy, C.; Nguyen, N.-K.; Guerardel, Y.; Chai, W.; Halary, F.; Renaudin-Autain, K.; Liu, Y.; Bressollette-Bodin, C.; Stehle, T.; McIlroy, D.  
Deposited on : 2022-07-20  
Resolution : 1.89 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4-5-2 with Phenix2.0  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 2.0  
EDS : 3.0  
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)  
CCP4 : 9.0.010 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.49

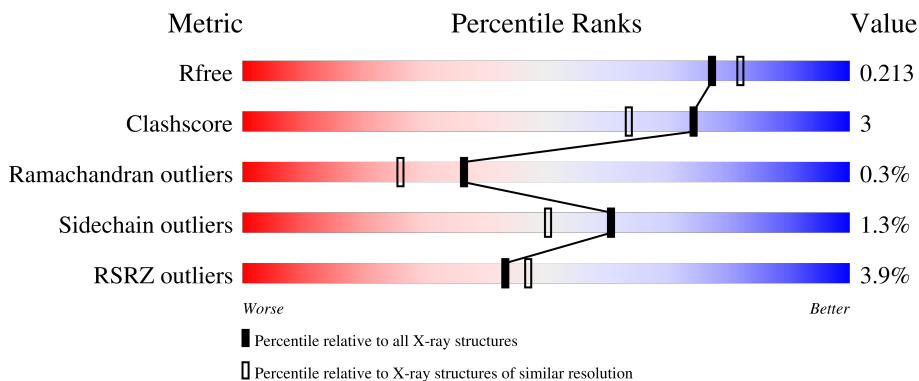
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.89 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	180053	1220 (1.88-1.88)
Clashscore	190562	1234 (1.88-1.88)
Ramachandran outliers	187476	1222 (1.88-1.88)
Sidechain outliers	187428	1222 (1.88-1.88)
RSRZ outliers	180081	1220 (1.88-1.88)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	AAA	271	 3% 89% 6% 5%
1	BBB	271	 4% 87% 8% . .
1	CCC	271	 4% 84% 9% 7%
1	DDD	271	 4% 87% 10% . .
1	EEE	271	 4% 88% 9% .

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 10521 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Major capsid protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	AAA	258	Total 1956	C 1232	N 336	O 376	S 12	0	0	0
1	BBB	261	Total 1971	C 1242	N 341	O 375	S 13	0	1	0
1	CCC	253	Total 1900	C 1197	N 327	O 363	S 13	0	2	0
1	DDD	266	Total 2003	C 1262	N 346	O 383	S 12	0	0	0
1	EEE	262	Total 2030	C 1276	N 350	O 392	S 12	1	8	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	72	ALA	GLU	engineered mutation	UNP P03088
AAA	104	SER	CYS	engineered mutation	UNP P03088
BBB	72	ALA	GLU	engineered mutation	UNP P03088
BBB	104	SER	CYS	engineered mutation	UNP P03088
CCC	72	ALA	GLU	engineered mutation	UNP P03088
CCC	104	SER	CYS	engineered mutation	UNP P03088
DDD	72	ALA	GLU	engineered mutation	UNP P03088
DDD	104	SER	CYS	engineered mutation	UNP P03088
EEE	72	ALA	GLU	engineered mutation	UNP P03088
EEE	104	SER	CYS	engineered mutation	UNP P03088

- Molecule 2 is GLYCEROL (CCD ID: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	AAA	1	Total C O 6 3 3	0	0
2	BBB	1	Total C O 6 3 3	0	0
2	CCC	1	Total C O 6 3 3	0	0
2	DDD	1	Total C O 6 3 3	0	0
2	EEE	1	Total C O 6 3 3	0	0

- Molecule 3 is CHLORIDE ION (CCD ID: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	AAA	1	Total Cl 1 1	0	0
3	BBB	1	Total Cl 1 1	0	0
3	CCC	1	Total Cl 1 1	0	0
3	DDD	1	Total Cl 1 1	0	0
3	EEE	1	Total Cl 1 1	0	0

- Molecule 4 is water.

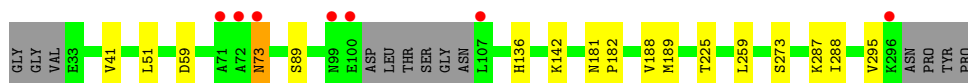
<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	AAA	130	Total 130	O 130	0	0
4	BBB	113	Total 113	O 113	0	0
4	CCC	121	Total 121	O 121	0	0
4	DDD	126	Total 126	O 126	0	0
4	EEE	136	Total 136	O 136	0	0

### 3 Residue-property plots [i](#)


These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

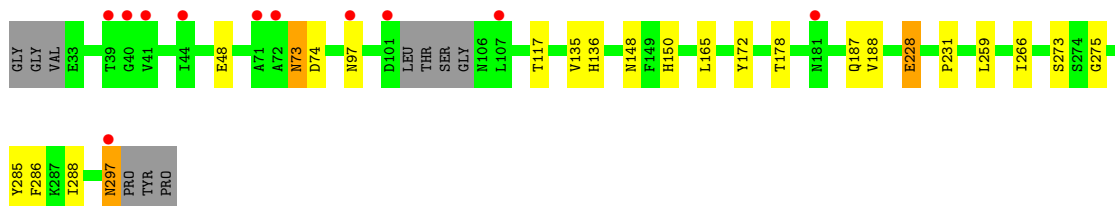
- Molecule 1: Major capsid protein VP1

Chain AAA: 




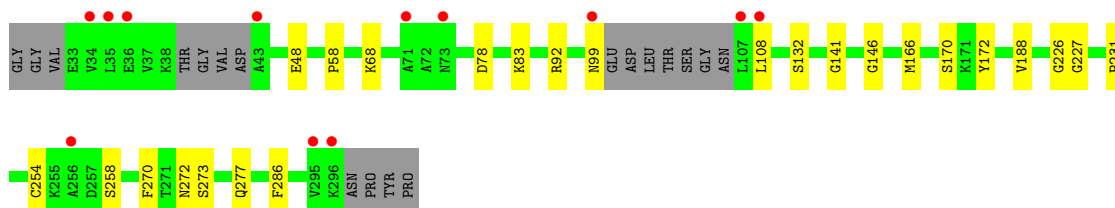
- Molecule 1: Major capsid protein VP1

Chain BBB: 



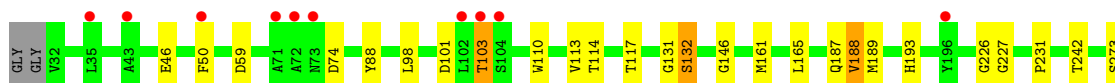
- Molecule 1: Major capsid protein VP1

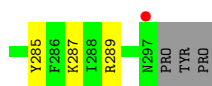
Chain CCC: 



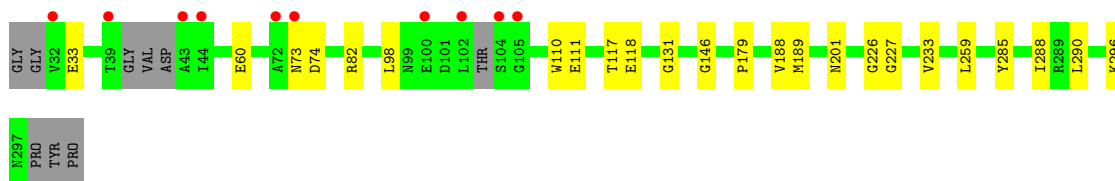
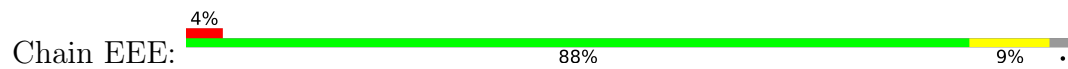
- Molecule 1: Major capsid protein VP1

Chain DDD: 





- Molecule 1: Major capsid protein VP1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	144.49Å 152.31Å 62.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.90 – 1.89 47.90 – 1.89	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.90-1.89) 99.9 (47.90-1.89)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.32 (at 1.88Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.174 , 0.211 0.182 , 0.213	Depositor DCC
$R_{free}$ test set	5587 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	28.6	Xtrriage
Anisotropy	0.153	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 31.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	10521	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.41% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CL, GOL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	AAA	1.02	0/2001	1.27	1/2725 (0.0%)
1	BBB	1.03	0/2016	1.25	1/2746 (0.0%)
1	CCC	1.06	1/1944 (0.1%)	1.27	2/2650 (0.1%)
1	DDD	1.06	0/2049	1.31	6/2794 (0.2%)
1	EEE	1.04	0/2083	1.22	1/2836 (0.0%)
All	All	1.04	1/10093 (0.0%)	1.26	11/13751 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	CCC	68	LYS	C-O	5.26	1.29	1.23

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	DDD	50	PHE	CA-C-O	-5.83	114.28	120.40
1	CCC	172	TYR	CB-CA-C	5.62	116.96	108.86
1	BBB	150	HIS	CB-CA-C	5.47	118.89	110.14
1	CCC	166	MET	N-CA-C	-5.44	105.43	111.36
1	DDD	59	ASP	CA-CB-CG	5.38	117.98	112.60

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within

the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	AAA	1956	0	1869	11	0
1	BBB	1971	0	1871	12	0
1	CCC	1900	0	1772	13	0
1	DDD	2003	0	1907	18	0
1	EEE	2030	0	1924	14	0
2	AAA	6	0	8	0	0
2	BBB	6	0	8	0	0
2	CCC	6	0	8	0	0
2	DDD	6	0	8	0	0
2	EEE	6	0	8	0	0
3	AAA	1	0	0	0	0
3	BBB	1	0	0	0	0
3	CCC	1	0	0	0	0
3	DDD	1	0	0	1	0
3	EEE	1	0	0	0	0
4	AAA	130	0	0	1	0
4	BBB	113	0	0	0	0
4	CCC	121	0	0	1	0
4	DDD	126	0	0	3	0
4	EEE	136	0	0	2	0
All	All	10521	0	9383	60	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 3.

The worst 5 of 60 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:CCC:99:ASN:HD21	1:CCC:108:LEU:H	1.32	0.76
1:DDD:132:SER:HB3	4:DDD:505:HOH:O	1.93	0.68
1:EEE:74[A]:ASP:OD1	4:EEE:501:HOH:O	2.10	0.67
1:BBB:297:ASN:N	1:BBB:297:ASN:OD1	2.30	0.63
1:DDD:131:GLY:N	1:EEE:73[B]:ASN:O	2.32	0.63

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	AAA	254/271 (94%)	244 (96%)	10 (4%)	0	100	100
1	BBB	258/271 (95%)	242 (94%)	14 (5%)	2 (1%)	16	5
1	CCC	249/271 (92%)	239 (96%)	9 (4%)	1 (0%)	30	18
1	DDD	264/271 (97%)	256 (97%)	7 (3%)	1 (0%)	30	18
1	EEE	264/271 (97%)	250 (95%)	14 (5%)	0	100	100
All	All	1289/1355 (95%)	1231 (96%)	54 (4%)	4 (0%)	36	26

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	BBB	188	VAL
1	CCC	188	VAL
1	DDD	188	VAL
1	BBB	74	ASP

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	AAA	209/231 (90%)	205 (98%)	4 (2%)	50	36
1	BBB	207/231 (90%)	202 (98%)	5 (2%)	43	27
1	CCC	196/231 (85%)	195 (100%)	1 (0%)	81	76
1	DDD	211/231 (91%)	209 (99%)	2 (1%)	70	63

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	EEE	214/231 (93%)	213 (100%)	1 (0%)	81	76
All	All	1037/1155 (90%)	1024 (99%)	13 (1%)	61	50

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	273	SER
1	BBB	297	ASN
1	EEE	179	PRO
1	DDD	132	SER
1	DDD	273	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 5 are monoatomic - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	GOL	EEE	401	-	5,5,5	0.13	0	5,5,5	0.42	0
2	GOL	BBB	501	-	5,5,5	0.15	0	5,5,5	0.43	0
2	GOL	AAA	501	-	5,5,5	0.14	0	5,5,5	0.42	0
2	GOL	CCC	401	-	5,5,5	0.18	0	5,5,5	0.52	0
2	GOL	DDD	401	-	5,5,5	0.16	0	5,5,5	0.44	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	EEE	401	-	-	2/4/4/4	-
2	GOL	BBB	501	-	-	0/4/4/4	-
2	GOL	AAA	501	-	-	1/4/4/4	-
2	GOL	CCC	401	-	-	2/4/4/4	-
2	GOL	DDD	401	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	CCC	401	GOL	C1-C2-C3-O3
2	EEE	401	GOL	C1-C2-C3-O3
2	EEE	401	GOL	O2-C2-C3-O3
2	DDD	401	GOL	O1-C1-C2-C3
2	DDD	401	GOL	C1-C2-C3-O3

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ > 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q < 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	AAA	258/271 (95%)	-0.07	7 (2%) 56 61	22, 30, 49, 67	0
1	BBB	261/271 (96%)	0.03	11 (4%) 40 44	15, 31, 57, 86	1 (0%)
1	CCC	253/271 (93%)	0.12	12 (4%) 36 38	14, 33, 58, 77	2 (0%)
1	DDD	266/271 (98%)	0.04	11 (4%) 41 45	20, 30, 55, 66	0
1	EEE	262/271 (96%)	-0.09	10 (3%) 44 48	15, 28, 48, 92	8 (3%)
All	All	1300/1355 (95%)	0.01	51 (3%) 43 47	14, 31, 55, 92	11 (0%)

The worst 5 of 51 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	EEE	72[A]	ALA	4.6
1	DDD	196	TYR	4.3
1	DDD	72	ALA	4.2
1	EEE	102	LEU	4.2
1	BBB	41	VAL	3.9

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

### 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum,

median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	GOL	AAA	501	6/6	0.83	0.21	50,62,68,73	0
2	GOL	DDD	401	6/6	0.83	0.17	43,57,65,67	0
2	GOL	EEE	401	6/6	0.83	0.17	38,57,57,63	0
2	GOL	CCC	401	6/6	0.86	0.15	42,50,51,58	0
2	GOL	BBB	501	6/6	0.89	0.13	43,53,56,58	0
3	CL	CCC	402	1/1	0.90	0.15	51,51,51,51	0
3	CL	DDD	402	1/1	0.95	0.09	45,45,45,45	0
3	CL	EEE	402	1/1	0.96	0.09	45,45,45,45	0
3	CL	BBB	502	1/1	0.97	0.08	41,41,41,41	0
3	CL	AAA	502	1/1	0.97	0.11	39,39,39,39	0

## 6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.